# DRUG TARGETS IN CANDIDA ALBICANS

The present invention is concerned with the identification of genes or functional fragments thereof from *Candida albicans* which are critical for growth and cell division and which genes may be used as selective drug targets to treat *Candida albicans* associated infections. Novel nucleic acid sequences from *Candida albicans* are also provided and which encode the polypeptides which are critical for growth of *Candida albicans*.

Opportunistic infections in immunocompromised hosts represent an increasingly common cause of mortality and morbidity. *Candida* species are among the most commonly identified fungal pathogens associated with such opportunistic infections, with *Candida albicans* being the most common species. Such fungal infections are thus problematical in, for example, AIDS populations in addition to normal healthy women where *Candida albicans* yeasts represent the most common cause of vulvovaginitis.

Although compounds do exist for treating such disorders, such as for example, amphotericin, these drugs are generally limited in their treatment because of their toxicity and side effects. Therefore, there exists a need for new compounds which may be used to treat *Candida* associated infections in addition to compounds which are selective in their action against *Candida albicans*.

Classical approaches for identifying anti-fungal compounds have relied almost exclusively on inhibition of fungal or yeast growth as an endpoint. Libraries of natural products, semi-synthetic, or synthetic chemicals are screened for their ability to kill or arrest growth of the target pathogen or a related nonpathogenic model organism. These tests are

cumbersome and provide no information about a compounds mechanism of action. The promising lead compounds that emerge from such screens must then be tested for possible host-toxicity and detailed mechanism of action studies must subsequently be conducted to identify the affected molecular target.

The present inventors have now identified a range of nucleic acid sequences form *Candida albicans* which encode polypeptides which are critical for its survival and growth. These sequences represent novel targets which can be incorporated into an assay to selectively identify compounds capable of inhibiting expression of such polypeptides and their potential use in alleviating diseases or conditions associates with *Candida albicans* infection.

Therefore, according to a first aspect of the invention there is provided a nucleic acid molecule encoding a polypeptide which is critical for survival and growth of the yeast *Candida albicans* and which nucleic acid molecule comprises any of the sequences of nucleotides in Sequence ID Numbers 1, 2, 3, 5, 10, 11, 12, 14, 16, 18, 20, 21, 23, 25, 27, 29, 31, 33, 37, 39, 41, 44, 45, 46, 49, 50, 52, 55, 57, 59, 61, 63, 65, 67, 70, 72, 74, 76, 78, 80, 81, 83, 85, 87, 89, 91, 93, 95, 97, 99, 101, 104, 106, 108, 110 and 113, or the sequences of nucleotides identified in Figures 9 to 13.

Whilst the molecules defined herein have been established as being critical for growth and metabolism of Candida albicans, for some of the molecules no apparent functionality has been assigned by virtue of the fact that no functionally related sequences in other prokaryotic or eukaryotic organism can be found in respective databases. Thus, advantageously these sequences may be species specific in which case they may be used be used as selective targets for treatment of diseases mediated by *Candida*

*Albicans* infection. Thus, in one aspct of the invention the nucleic acid molecules preferably comprise the sequences identified in sequence ID Nos 1, 2, 3, 5, 10, 11, 12, 14, 16, 17, 18, 46, 49, 50, 52, 55, 57, 59, 61, 63, 65, 87, 89, 91, 93, 95, 97, 99, 101, 104, 106, 108, and 110 and the corresponding polypeptide sequences identified in Table 1.

Some of sequences according to invention have been assigned a particular function. Nucleic acid molecules according to this aspect of the invention comprise any of the sequences as described in sequence ID Nos, 20, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 45, 65, 70, 72, 74, 76, 78, 80, 81, 83, 85 and 113 and the corresponding polypeptides identified in Table 1

Letters utilised in the nucleic acid sequences according to the invention to represent the genetic code and which are not recognisable as letters of the genetic code signify a position in the nucleic acid sequence where one or more of bases A, G, C or T can occupy the nucleotide position. Representative ambiguity codes used to identify the range of bases which can be used are as follows:

| | |
|---|---|
| M: | A or C |
| R: | A or G |
| W: | A or T |
| S: | C or G |
| Y: | C or T |
| K: | G or T |
| V: | A or C or G |
| H: | A or C or T |
| D: | A or G or T |
| B: | C or G or T |
| N: | G or A or T or C |

In one embodiment of the above identified aspects

of the invention the nucleic acid may comprise a mRNA molecule or alternatively a DNA and preferably a cDNA molecule.

Also provided by the present invention is a nucleic acid molecule capable of hybridising to the nucleic acid molecules according to the invention under high stringency conditions, such as for example, an antisense molecule.

Stringency of hybridisation as used herein refers to conditions under which polynucleic acids are stable. The stability of hybrids is reflected in the melting temperature (Tm) of the hybrids. Tm can be approximated by the formula:

$$81.5°C + 16.6\ (\log_{10}[Na^+] + 0.41\ (\%G\&C) - 600l/l$$

wherein l is the length of the hybrids in nucleotides. Tm decreases approximately by 1-1.5°C with every 1% decrease in sequence homology.

The nucleic acid capable of hybridising to nucleic acid molecules according to the invention will generally be at least 70%, preferably at least 80 or 90% and more preferably at least 95 to 97% homologous to the nucleotide sequences according to the invention.

The DNA molecules according to the invention may, advantageously, be included in a suitable expression vector to express polypeptides encoded therefrom in a suitable host.

The present invention also comprises within its scope proteins or polypeptides encoded by the nucleic acid molecules according to the invention or a functional equivalent, derivative or bioprecursor thereof.

Therefore, according to a further aspect of the invention there is provided a polypeptide which is critical for the growth and survival of Candida

albicans comprising an amino acid sequence of any of Sequence ID Numbers 4, 6 to 9, 13, 15, 19, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 43, 47, 48, 51, 53, 54, 56, 58, 60, 62, 64, 66, 68, 69, 71, 73, 75, 77, 79, 82, 84, 86, 88, 90, 92, 94, 96, 98, 100, 102, 103, 105, 107, 109, 111, 112, 114 or the sequences illiustrated in Figures 14 or 15.

An expression vector according to the invention includes a vector having a nucleic acid according to the invention operably linked to regulatory sequences, such as promoter regions, that are capable of effecting expression of said DNA fragments. The term "operably linked" refers to a juxta position wherein the components described are in a relationship permitting them to function in their intended manner. Such vectors may be transformed into a suitable host cell to provide for expression of a polypeptide according to the invention. Thus, in a further aspect, the invention provides a process for preparing polypeptides according to the invention which comprises cultivating a host cell, transformed or transfected with an expression vector as described above under conditions to provide for expression by the vector of a coding sequence encoding the polypeptides, and recovering the expressed polypeptides.

The vectors may be, for example, plasmid, virus or phage vectors provided with an origin of replication, optionally a promoter for the expression of said nucleotide and optionally a regulator of the promoter. The vectors may contain one or more selectable markers, such as, for example, ampicillin resistance.

Polynucleotides according to the invention may be inserted into the vectors described in an antisense orientation in order to provide for the production of antisense RNA. Antisense RNA or other antisense

nucleic acids may be produced by synthetic means.

In accordance with the present invention, a defined nucleic acid includes not only the identical nucleic acid but also any minor base variations including in particular, substitutions in bases which result in a synonymous codon (a different codon specifying the same amino acid residue) due to the degenerate code in conservative amino acid substitutions. The term "nucleic acid sequence" also includes the complementary sequence to any single stranded sequence given regarding base variations.

The present invention also advantageously provides nucleic acid sequences of at least approximately 10 contiguous nucleotides of a nucleic acid according to the invention and preferably from 10 to 50 nucleotides. These sequences may, advantageously be used as probes or primers to initiate replication, or the like. Such nucleic acid sequences may be produced according to techniques well known in the art, such as by recombinant or synthetic means. They may also be used in diagnostic kits or the like for detecting the presence of a nucleic acid according to the invention. These tests generally comprise contacting the probe with the sample under hybridising conditions and detecting for the presence of any duplex or triplex formation between the probe and any nucleic acid in the sample.

According to the present invention these probes may be anchored to a solid support. Preferably, they are present on an array so that multiple probes can simultaneously hybridize to a single biological sample. The probes can be spotted onto the array or synthesised *in situ* on the array. (See Lockhart *et al.*, Nature Biotechnology, vol. 14, December 1996 "Expression monitoring by hybridisation to high density oligonucleotide arrays". A single array can contain more than 100, 500 or even 1,000 different

probes in discrete locations.

Advantageously, the nucleic acid sequences, according to the invention may be produced using such recombinant or synthetic means, such as for example, using PCR cloning mechanisms which generally involve making a pair of primers, which may be from approximately 10 to 50 nucleotides to a region of the gene which is desired to be cloned, bringing the primers into contact with mRNA, cDNA, or genomic DNA from a human cell, performing a polymerase chain reaction under conditions which bring about amplification of the desired region, isolating the amplified region or fragment and recovering the amplified DNA. Generally, such techniques as defined herein are well known in the art, such as described in Sambrook *et al* (Molecular Cloning: a Laboratory Manual, 1989).

The nucleic acids or oligonucleotides according to the invention may carry a revealing label. Suitable labels include radioisotopes such as $^{32}P$ or $^{39}S$, enzyme labels or other protein labels such as biotin or fluorescent markers. such labels may be added to the nucleic acids or oligonucleotides of the invention and may be detected using known techniques *per se*.

The polypeptide or protein according to the invention includes all possible amino acid variants encoded by the nucleic acid molecule according to the invention including a polypeptide encoded by said molecule and having conservative amino acid changes. Polypeptides according to the invention further include variants of such sequences, including naturally occurring allelic variants which are substantially homologous to said polypeptides. In this context, substantial homology is regarded as a sequence which has at least 70%, preferably 80 or 90% amino acid homology with the polypeptides encoded by

the nucleic acid molecules according to the invention.

A nucleic acid which is particularly advantageous is one comprising the sequences of nucleotides according to Seq ID Nos 1 and 91 in which are specific to *Candida albicans* with no functionally related sequences in other prokaryotic or eukaryotic organism as yet identified from the respective genomic databases.

Nucleotide sequences according to the invention are particularly advantageous for selective therapeutic targets for treating *Candida albicans* associated infections. For example, an antisense nucleic acid capable of binding to the nucleic acid sequences according to the invention may be used to selectively inhibit expression of the corresponding polypeptides, leading to impaired growth of the *Candida albicans* with reductions of associated illnesses or diseases.

The nucleic acid molecule or the polypeptide according to the invention may be used as a medicament, or in the preparation of a medicament, for treating diseases or conditions associated with *Candida albicans* infection.

Advantageously, the nucleic acid molecule or the polypeptide according to the invention may be provided in a pharmaceutical composition together with a pharmaceutically acceptable carrier, diluent or excipient therefor.

Antibodies to the protein or polypeptide of the present invention may, advantageously, be prepared by techniques which are known in the art. For example, polyclonal antibodies may be prepared by inoculating a host animal, such as a mouse, with the polypeptide according to the invention or an epitope thereof and recovering immune serum. Monoclonal antibodies may be prepared according to known techniques such as described by Kohler R. and Milstein C., Nature

(1975)256, 495-497.

Antibodies according to the invention may also be used in a method of detecting for the presence of a polypeptide according to the invention, which method comprises reacting the antibody with a sample and identifying any protein bound to said antibody. A kit may also be provided for performing said method which comprises an antibody according to the invention and means for reacting the antibody with said sample.

Proteins which interact with the polypeptide of the invention may be identified by investigating protein-protein interactions using the two-hybrid vector system first proposed by Chien *et al* (1991).

This technique is based on functional reconstitution *in vivo* of a transcription factor which activates a reporter gene. More particularly the technique comprises providing an appropriate host cell with a DNA construct comprising a reporter gene under the control of a promoter regulated by a transcription factor having a DNA binding domain and an activating domain, expressing in the host cell a first hybrid DNA sequence encoding a first fusion of a fragment or all of a nucleic acid sequence according to the invention and either said DNA binding domain or said activating domain of the transcription factor, expressing in the host at least one second hybrid DNA sequence, such as a library or the like, encoding putative binding proteins to be investigated together with the DNA binding or activating domain of the transcription factor which is not incorporated in the first fusion; detecting any binding of the proteins to be investigated with a protein according to the invention by detecting for the presence of any reporter gene product in the host cell; optionally isolating second hybrid DNA sequences encoding the binding protein.

An example of such a technique utilises the GAL4

protein in yeast. GAL4 is a transcriptional activator of galactose metabolism in yeast and has a separate domain for binding to activators upstream of the galactose metabolising genes as well as a protein binding domain. Nucleotide vectors may be constructed, one of which comprises the nucleotide residues encoding the DNA binding domain of GAL4. These binding domain residues may be fused to a known protein encoding sequence, such as for example the nucleic acids according to the invention. The other vector comprises the residues encoding the protein binding domain of GAL4. These residues are fused to residues encoding a test protein. Any interaction between polypeptides encoded by the nucleic acid according to the invention and the protein to be tested leads to transcriptional activation of a reporter molecule in a GAL-4 transcription deficient yeast cell into which the vectors have been transformed. Preferably, a reporter molecule such as β-galactosidase is activated upon restoration of transcription of the yeast galactose metabolism genes.

Further provided by the present invention is one or more *Candida albicans* cells comprising an induced mutation in the DNA sequence encoding the polypeptide according to the invention.

A further aspect of the invention provides a method of identifying compounds which selectively inhibit or interfere with the expression, or the functionality of polypeptides expressed from the nucleotides sequences according to the invention or the metabolic pathways in which these polypeptides are involved and which are critical for growth and survival of *Candida albicans*, which method comprises (a) contacting a compound to be tested with one or more *Candida albicans* cells having a mutation in a nucleic acid molecule according to the invention which mutation results in overexpression or underexpression

of said polypeptides in addition to one or more wild type *Candida* cells, (b) monitoring the growth and/or activity of said mutated cell compared to said wild type wherein differential growth or activity of said one or more mutated *Candida* cells provides an indication of selective action of said compound on said polypeptide or another polypeptide in the same or a parallel pathway.

Compounds identifiable or identified using the method according to the invention, may advantageously be used as a medicament, or in the preparation of a medicament to treat diseases or conditions associated with *Candida albicans* infection. These compounds may also advantageously be included in a pharmaceutical composition together with a pharmaceutically acceptable carrier, diluent or excipient therefor.

A further aspect of the invention provides a method of identifying DNA sequences from a cell or organism which DNA encodes polypeptides which are critical for growth or survival, which method comprises (a) preparing a cDNA or genomic library from said cell or organism in a suitable expression vector which vector is such that it can either integrate into the genome in said cell or that it permits transcription of antisense RNA from the nucleotide sequences in said cDNA or genomic library, (b) selecting transformants exhibiting impaired growth and determining the nucleotide sequence of the cDNA or genomic sequence from the library included in the vector from said transformant. Preferably, the cell or organism may be any yeast or filamentous fungi, such as for example, *Saccharomyces cervisiae*, *Saccharomyces pombe* or *Candida albicans*.

A further aspect of the invention provides a pharmaceutical composition comprising a compound according to the invention together with a

pharmaceutically acceptable carrier, diluent or excipient therefor.

The present invention may be more clearly understood with reference to the accompanying example, which is purely exemplary, with reference to the accompanying drawings wherein:

Figure 1: is an illustration of A) Intergration of the antisense library plasmid (here shown as a linear fragment) at a site (eg. *GAL1* promoter region) within the genome which is non-homologous to the insert DNA. As a result the *GAL1*p region is duplicated and antisense RNA can be formed from GENE X upon induction of *GAL1*p, and B) Intergration due to homologous recombination of the gene insert (GENE X) of an antisense library clone (here shown as a linear fragment) with the homologous gene (gene x) within the *Candida* genome. As a result this gene is duplicated. The first copy of the gene geNE X, is flanked by upstream its endogenous promoter and downstream, oppositely-oriented, the *GAL1* promoter resulting in a so-called "collision construct". Antisense RNA can be formed from GENE X upon induction of *GAL1*p. The second copy of the gene, GEne X, is devoid of a promoter and will not be transcribed.

Figure 2: is an illustration of the vectors used for the preparation of a cDNA antisense library, p*GAL1*PNiST-1, (left) and a genomic library, p*GAL1*PNiST-1 (right).

Figure 3: Growth curves in S-glucose and S-galactose medium of respectively the wild type CAI-4 strain and two transformants (clone 36 and 38) showing antisense induced reduction in growth and overall impaired growth, respectively. Growth curves in S-glucose+maltose and S-galactose+maltose medium of respectively the wild type CAI-4 strain and transformants resulting from antisense library transformation.

Figure 4: is an illustration of promoter activity of the *C. albicans GAL1* promoter in the absence and presence of maltose as a carbon source.

Figures 5: is a Northern blot analysis of *C. albicans* mRNA in wild type and clone 36 using a *SAM2* and a *TEF3* specific probe.

Figures 6: is A) a Northern blot analysis of sequences of *C. albicans* mRNA in wild type and clone 38 using a *RNR1* and an *ACT1* specific probe; and B) Real Time Quantitative PCR

on *C. albicans* mRNA in wild type and clone 38 using a *RNR1* and *ACT1* specific fluorogenic probe.

Figure 7: is a nucleotide sequence of plasmid pGAL1PNiST-1.

Figure 8: is a nucleotide sequence of plasmid pGAL1PSiST-1.

Figure 9: is a nucleotide sequence of clone 38 which has been assigned RNR1 functionally.

Figure 10: is a nucleotide sequence of clone 113g4.

Figure 11: is a nucleotide sequence of clone 207g4

Figure 12: is a nucleotide sequence of clone 66g4.

Figure 13: is a nucleotide sequence of clone 36 which has been assigned Sam2 functionally.

Figure 14: is an amino acid sequence of clone 38.

Figure 15: is an amino acid sequence of clone 36.

Figures 16 to 70 are growth curves of Candida albicans showing antisense induced reduction in growth by inhibition of molecules according to the invention.

Figure 71 is an illustration of the results obtained in a Northern blot analysis of a FAL1 single allele knock-out.

Lane 1: RNA MWM I (Boerhinger Mannheim)
Lane 2: WT + gal + mal + LiAc
Lane 3: FAL1 + gal + mal + LiAc
Lane 4: RNA MWM I DIG labelled (Boerhinger Mannheim)

Lower FAL1 transcript levels are observed in the FAL1 single allele knock-out strain compared to the wild type strain.

Figure 72 is an illustration of growth curves obtained of a FAL1 single allele knock-out strain and wild type. The FAL1 single allele knock-out grows equal to the wild type. However, it is significantly more resistant to Hygromycin B.

Figure 73 is an illustration of the results of a Northern blot obtained using a RNR1 single allele knock-out. Correct and single integration of RNR1 disruption cassette was confirmed by PCR and Southern blot analysis. Lower RNR1 transcript

levels are observed in the RNR1 single allele knock-out strain compared to the wild type strain. This result was confirmed by quantitative PCR (QT-PCR).

Figure 74: is an illustration of the growth curves obtained from various RNR1 single allele knock-out strains. The RNR1 single allele knock-out shows an extended LAG phase compared to wild type.

Figure 75 is an illustration of the growth curves obtained from a single SAM2 knock-out clone against wild type. Inoculum for SAM2 was somewhat higher; at equal inocula, growth of SAM2 single allele knock-out is slightly higher.

Figure 76 is an illustration of growth curves obtained in a RHO1 single allele knock-out. Growth of the RHO1 single allele knock-out is impaired compared to wild type growth.

Figure 77 is an illustration of the results obtained from an RNA level analysis for a MEG1 single allele knock-out. MEG1 expression was decreased more than 14 fold in the MEG1 single allele knock-out compared to the WT.

Figure 78 is an illustration of the results

obtained in a growth curve analysis of MEG1 single allele knock-out compared to wild type. Inoculum for SAM2 was somewhat higher; at equal inocula growth of SAM2 single allele knock-out is slightly lower.

Figure 79 is an illustration of the results obtained from an RNA level QT-PCR analysis on a MAA single allele knock-out. Correct and single integration of MAA disruption cassette was confirmed by both PCT and Southern blot analysis. MAA expression was decreased two fold in MAA knock-out compared to wild type.

Figure 80 is an illustration of the growth curves obtained in a single allele knock-out MAA clone compared to wild type. Inoculum for MAA was somewhat higher; at equal inocula growth of MAA single allele knock-out is slightly slower.

Figure 81 is an illustration of the results obtained from a RNA level QT-PCR analysis on a RPL27 single allele knock-out. RPL27 expression was decreased more than three fold in the RPC27 knock-out compared to wild type.

Figure 82 is an illustration of the growth curves obtained from a RPL27

single allele knock-out clone compared to wild type. The RPL27 single allele knock-out strain grows equally to the wild type strain.

**Example**

Identification of novel drug targets in C. albicans by anti-sense and disruptive integration

The principle of the approach is based on the fact that when a particular C. albicans mRNA is inhibited by producing the complementary anti-sense RNA, the corresponding protein will decrease. If this protein is critical for growth or survival, the cell producing the anti-sense RNA will grow more slowly or will die.

Since anti-sense inhibition occurs at mRNA level, the gene copy number is irrelevant, thus allowing applications of the strategy even in diploid organisms.

Anti-sense RNA is endogenously produced from an integrative or episomal plasmid with an inducible promoter; induction of the promoter leads to the production of a RNA encoded by the insert of the plasmid. This insert will differ from one plasmid to another in the library. The inserts will be derived from genomic DNA fragments or from cDNA to cover-to the extent possible- the entire genome.

The vector is a proprietary vector allowing integration by homologous recombination at either the homologous insert or promoter sequence in the Candida genome. After introducing plasmids from cDNA or genomic libraries into C. albicans, transformants are screened for impaired growth after promoter (& thus anti-sense) induction in the presence of lithium acetate. Lithium acetate prolongs the G1 phase and

thus allows anti-sense to act during a prolonged period of time during the cell cycle. Transformants which show impaired growth in both induced and non-induced media, thus showing a growth defect due to integrative disruption, are selected as well.

Transformants showing impaired growth are supposed to contain plasmids which product anti-sense RNA or mRNAs critical for growth or survival. Growth is monitored by measuring growth-curves over a period of time in a device (Bioscreen Analyzer, Labsystems) which allows simultaneous measurement of growth-curves of 200 transformants.

Subsequently plasmids can be recovered from the transformants and the sequence of their inserts determined, thus revealing which mRNA they inhibit. In order to be able to recover the genomic or cDNA insert which has integrated into the Candida genome, genomic DNA is isolated, cut with an enzyme which cuts only once into the library vector (and estimated approx. every 4096 bp in the genome) and relegated. PCR with primers flanking in the insert will yield (Partial) genomic or cDNA inserts as PCR fragments which can directly be sequenced. This PCR analysis (on ligation reaction) will also show us how many integrations occurred. Alternatively the ligation reaction is transformed to E. coli and PCR analysis is performed on colonies or on plasmid DNA derived thereof.

This method is employed for a genome wide search for novel C. albicans genes which are important for growth or survival.

## MATERIALS AND METHODS

### Construction of pGal1NIST-1

pGAL1PNiST-1 (integrative antisense SfiI-NotI vector)

was constructed as described by Logghe et al., submitted.

**Construction of pGAL1PSiST-1**

The vector pGAL1PSiST-1 (integrative SfiI-SfiI vector) was created for cloning the small genomic DNA fragments behind the GAL1 promoter. The only difference with pGAL1PNiST-1 is that the hIFNb insert fragment in pGAL1PSiST-1 is flanked by two SfiI sites instead of a SfiI and a NotI site as in pGAL1PNiST-1. To construct pGAL1PSiST-1 the EcoRI-HindIII fragment, containing hIFNb flanked by a SfiI and a NotI site, of pMAL2pHiET-3 (Logghe M., unpublished) was exchanged by the EcoRI-HindIII fragment, containing hIFNb flanked by two SfiI sites, from YCp50S-S (an E. coli / S. cerevisiae shuttle vector derived from the plasmid YCp50, which is deposited in the ATCC collection (number 37419; Thrash et al., 1985); an EcoRI-HindIII fragment, containing the gene hIFNb, which is flanked by two SfiI sites, was inserted in YCp50, creating YCp50S-S), resulting into plasmid pMAL2PSiST-1. The MAL2 promoter from pMAL2PSiST-1 (by a NaeI-FspI digest) was further replaced by the GAL1 promoter from pGAL1PNiST-1 (via a XhoI-SalI digest), creating the vector pGAL1PSiST-1.

**Preparation of C. albicans genomic library**

A C. albicans genomic DNA library with small DNA fragments was prepared for integrative disruption. Genomic DNA of C. albicans B2630 (ATCC No. 44858) was isolated following a modified protocol of Blin and Stafford (1976). To obtain enrichment for genomic DNA fragments of the desired size, the genomic DNA was partially digested. Enrichment of small DNA fragments

was obtained with 70 units of AluI on 10 mg of genomic DNA for 20 min. T4 DNA polymerase (Boehringer) and dNTPs (Boehringer) were added to polish the DNA ends. After extraction with phenol-chloroform the digest was size-fractionated on an agarose gel. The genomic DNA fragments with a length of 0.5 to 1.25 kb were eluted from the gel by centrifugal filtration (Zhu et al., 1985). SfiI adaptors (5' GTTGGCCTTTT) were attached to the DNA ends (blunt) to facilitate cloning of the fragments into the vector. After ligation of these adaptors to the DNA fragments a second size-fractionation was performed on an agarose gel. The small genomic DNA fragments were cloned upstream of the GAL1 promoter in the vector pGAL1PSiST-1. Qiagen-purified pGAL1PSiST-1 plasmid DNA was digested with SfiI and the largest vector fragment eluted from the gel by centrifugal filtration (Zhu et al., 1985). The ligation mix was electroporated to MC1061 (...) E. coli cells.

**C. albicans cDNA library**

Total RNA was extracted from C. albicans strain B2630 grown on respectively minimal (SD) and rich (YPD) medium as described by Sambrook et al. (1989). mRNA was prepared from total RNA using the Invitrogen Fast Track procedure. First strand cDNA was synthesised with Superscript Reverse Transcriptase (BRL) and with an oligo dT-NotI Primer adapter. After second strand synthesis, cDNA was polished with Klenow enzyme and purified over a Sephacryl S-400 spin column. Phosphorylated SfiI adapters were then ligated to the cDNA, followed by digestion with the NotI restriction enzyme. The SfiI/NotI cDNA was purified and sized on a Biogel column A150M. cDNA was ligated in a NotI/SfiI opened pGAL1PNiST-1 vector.

**Transformation of C. albicans**

C. albicans CAI-4 (URA3::imm434/URA3::imm434) was kindly provided by Dr. William Fonzi, Georgetown University (Fonzi and Irwin, 1993). CAI-4 was transformed with above described cDNA library or genomic library using a modified spheroplast method (Logghe M., submitted). Cells were plated on minimal medium supplemented with glucose and sorbitol (SD (0.67% Yeast Nitrogen base w/o amino acids + 2% glucose), 1 M sorbitol) plates using 0.4 cm glass-pearls (Glaverbel, Belgium) and incubated for 2-3 days at 30°C.

**Screening for mutants**

Starter cultures were set up by inoculating each colony in 1 ml SD medium and incubating overnight at 30°C and 300 rpm. Cell densities were determined using a Coulter counter (Coulter Z1; Coulter electronics limited). 250.000 cells/ml were inoculated in SD medium for a total volume of 1ml and cultures were incubated for 24 hours at 30°C and 300 rpm. Cultures were washed in minimal medium without glucose (S) and the pellet resuspended in 650 ml S medium. 8 µl of this culture was used for inoculating 400 µl cultures in a Honeywell-100 plate (Bioscreen analyzer, Labsystems). Each transformant was grown for three days in S medium containing 50 mM LiAc; pH 6.0, with 2% glucose +/- 2% maltose or 2% galactose +/- 2% maltose respectively while shaking (high intensity) every 3 minutes for 20 seconds. Optical densities were measured every hour and growth curves were generated automatically (Bioscreen analyzer; Labsystems).

**Construction of LAC4/ pGAL1PNiST-1**

pGAL1PNiST-1 vector was cut with StuI in order to release the stuffer fragment and subsequently dephosphorylated (CIP, Boehringer). Plasmid pRS1004, obtained from J. Ernst (University of Duesseldorf, Germany), was cut with PvuII/XbaI in order to release the K. lactis ß-galactosidase (EC 3.2.1.23; LAC4) reporter gene and Klenow-treated. The LAC4 PvuII/XbaI blunted reporter gene fragment from pRS1004 was ligated into StuI opened pGAL1PNiST-1 resulting in the integrative plasmid LAC4/ pGAL1PNiST-1

**Measurement of GAL1 promoter activity**

C. albicans strain CAI-4 was transformed with LAC4/pGAL1pNiST-1 using the modified spheroplast method (Logghe et al., submitted). Resulting transformants were grown in 5 ml of respectively non-induction (SD +/- maltose) and induction (S+ galactose +/- maltose) medium and further processed as described by Leuker et al. (1997).

**Isolation of genomic or cDNA inserts**

Potentially interesting transformants were grown in 1.5 ml SD overnight. Genomic DNA was isolated using the Nucleon MI Yeast kit (Amersham) and the concentration of genomic DNA was estimated by analyzing a sample on a 0.7% agarose gel in 0.5x TBE and comparison to a known standard molecular weight marker. 20 ng of genomic DNA was digested for three hours with an enzyme that cuts uniquely in the library vector (SacI for the genomic library; PstI for the cDNA library), treated with RNAse A (Boehringer) and incubated for 20 minutes at 65°C to inactivate the

enzyme. Samples were phenol/chloroform extracted twice and precipitated using NaOAc/ethanol. The resulting pellet was resuspended in 500 µl ligation mixture (1 x ligation buffer and T4 DNA ligase; both from Boehringer) and incubated overnight at 16°C. After denaturation (10 min 65°C), purification (phenol/chloroform extraction) and precipitation (NaOAc/ethanol) the pellet was resuspended in 10 µl MilliQ (Millipore) water.

Inverse PCR was performed on 1 µl of the precipitated ligation reaction using library vector specific primers (Figure 1) (3pGALSistPCR: 5' GAG-GGC-GTG-AAT-GTA-AGC-GTG 3' and 5pGALNistPCR: 5'GAG-TTA-TAC-CCT-GCA-GCT-CGA-C 3' for the genomic library; 3pGALNistPCR: 5' TGA-GCA-GCT-CGC-CGT-CGC-GC 3' and 5pGALNistPCR for the cDNA library; all primers from Eurogentec) for 30 cycles each consisting of (a) 1 min at 95 °C, (b) 1 min at 61 (or 57 °C for the cDNA library primers), and (c) 3 min at 72 °C. In the reaction mixture 2.5 units of Taq polymerase (Boehringer) with TaqStart antibody (Clontech) (1:1) were used, and the final concentrations were 0.2 µM of each primer, 3 mM $MgCl_2$ (Perkin Elmer Cetus) and 200 µM dNTPs (Perkin Elmer Cetus). All PCR reactions were performed in a Robocycler (Stratagene).

PCR analysis is also performed on genomic DNA isolated from the transformants using primers 3pGALSistPCR and 5pGALNistPCR for the genomic library transformants and using primers oligo23': 5' TGC-AGC-TCG-ACC-TCG-AGG 3' and oligo25: 5' GCG-TGA-ATG-TAA-GCG-TGA-C 3' ($T_{hybr}$ = 53 °C) for the cDNA library transformants.

Resulting PCR products were purified using the PCR purification kit (Qiagen) and were quantified by comparison of band intensity with the intensity of DNA marker bands on a ethidium bromide stained agarose gel.

**Sequence determination**

The amount of PCR product (expressed in ng) put in the sequencing reaction is calculated as the length of the PCR product in basepairs divided by 10. DNA sequencing reactions were performed using the ABI Prism BigDye Terminator Cycle Sequencing Ready Reaction Kit according to the instructions of the manufacturer (PE Applied Biosystems, Foster City, CA) except for the following modifications. The total reaction volume was reduced to 15 µl. Reaction volumes of individual reagents were changed accordingly. The 6.0 µl Terminator Ready Reaction Mix was replaced by a mixture of 3.0 µl Terminator Ready Reaction Mix + 3.0 µl Half Term (GENPAK Limited, Brighton, UK). After cycle sequencing, reaction mixtures were purified over Sephadex G50 columns prepared on Multiscreen HV opaque Microtiter plates (Millipore, Molsheim, Fr) and were dried in a speedVac. Reaction products were resuspended in 3 µl loading buffer. Following denaturation for 2 min at 95°C, 1 µl of sample was applied on a 5% Long Ranger Gel (36 cm well-to-read) prepared from Singel Packs according to the supplier's instructions (FMC BioProducts, Rockland, ME). Samples were run for 7 hours 2X run on a ABI 377XL DNA sequencer. Data collection version 2.0 and Sequence analysis version 3.0 (for basecalling) software packages are from PE Applied Biosystems.

**Sequence analysis**

Nucleotide sequences were imported in the VectorNTI software package (InforMax Inc, North Bethesda, MD, USA), and the vector and insert regions of the sequences were identified. Sequence similarity searches against public and commercial sequence databases were performed with the BLAST software package (Altschul et al., 1990) version 1.4. Both the

original nucleotide sequence and the six-frame conceptual translations of the insert region were used as query sequences. The used public databases were the EMBL nucleotide sequence database (Stoesser et al., 1998), the SWISS-PROT protein sequence database and its supplement TrEMBL (Bairoch and Apweiler, 1998), and the ALCES Candida albicans sequence database (Stanford University, University of Minnesota). The commercial sequence databases used were the LifeSeq® human and PathoSeq™ microbial genomic databases (Incyte Pharmaceuticals Inc., Palo Alto, CA, USA), and the GENESEQ patent sequence database (Derwent, London, UK). Three major results were obtained on the basis of the sequence similarity searches: function, novelty, and specificity. A putative function was deduced on the basis of the similarity with sequences with a known function, the novelty was based on the absence or presence of the sequences in public databases, and the specificity was based on the similarity with vertebrate homologues.

The 5' UTR region of the SAM2 gene was analysed using the "Findpatterns" algorithm of the Genetics Computer Group (GCG) software package (University of Wisconsin, USA).

**Northern blot analysis**

Cells were grown to $OD_{600}$ ~ 1.0 and total RNA was prepared using the RNeasy midi kit (Qiagen) according to the manufacturer's instructions. RNA concentrations were determined spectrophotometrically by measuring optical densities at 260 nm in a UV-1601 UV-visible spectrophotometer (Shimadzu) and 5 µg of each sample was resolved onto a 1% formaldehyde gel and run in 1 x formaldehyde gel running buffer (5prime-3prime) at 3.5 V/cm. RNA was stained for 20 minutes using SYBR Green II stain (Molecular probes) 1/10000 diluted in 1x formaldehyde gel running buffer (5prime-3prime) and

subsequently transferred to Hybond-N+ nylon membrane (Amersham) by overnight capillary blotting in 20 x SSC. DIG-labeled probes were prepared using DIG-dUTP (Boehringer Mannheim) at a 1:3 or 1:6 dTTP:DIG-dUTP ratio, 10 pg of template plasmid DNA, 1x PCR buffer II (Perkin Elmer Cetus), 10 µM of each primer (Eurogentec), 0.2 mM of dATP, dCTP and dGTP (Perkin Elmer Cetus), 2.5 mM $MgCl_2$ (Perkin Elmer Cetus), 5% DMSO and 1.25 units Taq polymerase (Boehringer). The membrane was prehybridized at 50°C (DNA probes) or at 68°C (RNA probes) in DIG Easy Hyb (Boehringer Mannheim) for minimum 1 hour. Hybridization was performed using 1 µl PCR reaction product (= 1/50 of the total volume)/ml DIG Easy Hyb. The probes were denatured by heating the PCR reaction for 10 minutes at 96°C, then quick-chilling on ice. The probe was kept on ice for 5 minutes, centrifuged briefly and diluted in pre-warmed DIG Easy Hyb solution. The entire probe solution was filtered through a 0.45 µm filter (Millex HV, Millipore) prior to use. Hybridizations were carried out overnight. Post-hybridization, membranes were washed twice 15 minutes with 2x SSC/0.1% SDS at room temperature and twice 15 minutes with 0.1x SSC/0.1% SDS at 68°C. Detection was performed using the DIG Wash and Block Buffer Set as described by the manufacturer (Boehringer Mannheim Mannheim) and the blot was exposed to Kodak XAR-5 film for 1 hour at ambient temperature.

Real time quantitation of mRNA transcript

PCR quantitations using specific primers and probes were performed according to the TaqMan procedure (Livak et al., 1995; Orlando et al., 1998) using the ABI Prism 7700 sequence detector (Applied Biosystems). Primers and probes for ACT1 (b-actin) and RNR1 genes were designed using the PrimerExpress software system

(Perkin Elmer Cetus). Cells were grown to $OD_{600}$ ~ 1.0 and total RNA was prepared using the RNeasy midi kit (Qiagen) according to the manufacturer's instructions. All RNA samples were DNaseI (Boehringer-Mannheim, RNAse-free)-treated at 20 U/µg in 50 µl solution for 40 min at ambient temperature, phenol/chloroform-extracted and precipitated. Pellets were dissolved in 20 ml MilliQ water (Millipore) and RNA concentrations were determined spectrophoto-metrically. First-strand cDNA synthesis was performed in a final volume of 20 µl containing 1x Superscript RT buffer (Life Technologies), 10 mM DTT, 125 µM of each dNTP, 50 µM hexamer primers (Life Technologies) and 1 mg RNA. Mixtures were incubated for 10 min. at ambient temperature and 1 µl was removed and diluted 1:4 for the non-amplification control (NAC); 20 U Superscript reverse transcriptase (Life Technologies) was added and the reaction was incubated for 1 hour at 42 °C. The enzyme was inactivated for 10 min at 70°C. PCR reactions were set up in triplicate for all genes and contained 5 ml PCR buffer A, 4 mM $MgCl_2$, 200 µM each of dATP, dGTP, dCTP and 400 µM dUTP, 250 nM fluorogenic probe (for RNR1: 5' TGA-TCT-CAA-AAA-GTG-CTG-GAG-GAA-TCG-GT 3'), 0.5 U UNG, 1.25 U AmpliTaq Gold, 16.75 ml $H_2O$, 300 nM of appropriate FORWARD (for RNR1: 5' CGA-CAC-TTT-GAA-ATC-GTG-TGC-T 3') and REVERSE (for RNR1: 5' GCA-CCG-GTA-GAA-CGA-ATG-TTG 3') PCR primers, 1 ml of the RT reaction mixture.

For the NAC, 1 µl of the 1:4 diluted RTase-negative sample was added while 1 µl of $H_2O$ was added to each non-template control sample. The ABI PRISM 7700 was run for 50 cycles of 15 s at 95°C, 1 min at 60°C. These cycles were preceded by 5 min at 50°C (UNG activation) and 10 min at 95°C (UNG inactivation and DNA denaturation).

Data were analyzed using the ABI PRISM 7700 software package. Data were normalized according to ACT1 $C_T$

values.

**Library screening**

Using primers 5pGalNistPCR and 3pGalNistPCR, a 0.6 kb region of the C. albicans SAM2 gene was PCR-amplified from a SAM2/pGAL1pNiST-1 construct isolated from clone 36 and labeled with [$^{32}$P]dCTP using the Multiprime™ random-primed labeling system (Amersham). C. albicans genomic DNA isolated from strain B2630 was partially digested with Sau3AI, resolved on a 0.7% agarose gel and the region of the gel with the fragment size of interest (10-23kb) was cut out and DNA was eluted from the gel with Sephaglass Band Prep kit (Pharmacia). A C. albicans library in pYCP50 was prepared by ligating these fragments into a BamHI cut and dephosphorylated pYCP50 vector in a 1:2 molar ratio vector to insert. The titer (#colonies/µg DNA) was determined by transforming a fraction of the library to E. coli. Five genome equivalents were plated out and filter-lifts were prepared as described (Sambrook et al., 1989). Duplicate nylon filters were pre-washed for 2 hours at 42°C in 50 mM Tris, 1M NaCl, 0.1% SDS, 1 mM EDTA to reduce background hybridization. The filters were subsequently hybridized at 42°C overnight in 5x SSPE, 50% formamide, 5x Denhardt's solution, 0.1% SDS, 100 µg/ml denatured salmon sperm DNA and $10^6$ cpm/ml of denatured probe. Filters were then washed in 2x SSC, 0.5 % SDS for 1 hour at room temperature and for 1 hour at 50°C. A few intense autoradiographic spots were found and the corresponding colonies were selected for plasmid preparation. Candidate clones were digested with a panel of restriction enzymes, resolved on a 0.7 % agarose gel, stained with ethidiumbromide and transferred to nylon membrane by vacuum-blotting. The blot was probed under the same conditions as the genomic library. A 1.1 kb HpaI

fragment covering the entire hybridizing segment was subcloned into pCR-Blunt (Invitrogen)

**Screening for compounds modulating expression of polypeptides critical for growth and survival of C. albicans**

The method proposed is based on observations (Sandbaken et al., 1990; Hinnebusch and Liebman 1991; Ribogene PCT WO 95/11969, 1995) suggesting that underexpression or overexpression of any component of a process (e.g. translation) could lead to altered sensitivity to an inhibitor of a relevant step in that process. Such an inhibitor should be more potent against a cell limited by a deficiency in the macromolecule catalysing that step and/or less potent against a cell containing an excess of that macromolecule, as compared to the wild type (WT) cell.

Mutant yeast strains, for example, have shown that some steps of translation are sensitive to the stoichiometry of macromolecules involved. (Sandbaken et al.). Such strains are more sensitive to compounds which specifically perturb translation (by acting on a component that participates in translation) but are equally sensitive to compounds with other mechanisms of action.

This method thus not only provides a means to identify whether a test compound perturbs a certain process but also an indication of the site at which it exerts its effect. The component which is present in altered form or amount in a cell whose growth is affected by a test compound is potentially the site of action of the test compound.

The assay to be set up involves measurement of growth of an isogenic strain which has been modified only in a certain specific allele, relative to a wild type (WT) C. albicans strain, in the presence of R-

compounds. Strains can be ones in which the expression of a specific essential protein is impaired upon induction of anti-sense or strains which carry disruptions in an essential gene. An in silico approach to finding novel essential genes in C. albicans will be performed. A number of essential genes identified in this way will be disrupted (in one allele) and the resulting strains can be used for comparative growth screening.

Assay for High Throughput screening for drugs

35 $\mu$l minimal medium (S medium + 2% galactose + 2% maltose) is transferred in a transparent flat-bottomed 96 well plate using an automated pipetting system (Multidrop, Labsystems). A 96-channel pipettor (Hydra, Robbins Scientific) transfers 2.5 $\mu$l of R-compound at $10^{-3}$ M in DMSO from a stock plate into the assay plate.

The selected C. albicans strains (mutant and parent (CAI-4) strain) are stored as glycerol stocks (15%) at -70°C. The strains are streaked out on selective plates (SD medium) and incubated for two days at 30°C. For the parent strain, CAI-4, the medium is always supplemented with 20 $\mu$g/ml uridine. A single colony is scooped up and resuspended in 1 ml minimal medium (S medium + 2% galactose + 2% maltose). Cells are incubated at 30°C for 8 hours while shaking at 250 rpm. A 10 ml culture is inoculated at 250.000 cells/ml. Cultures are incubated at 30°C for 24 hours while shaking at 250 rpm. Cells are counted in Coulter counter and the final culture (S medium + 2% galactose + 2% maltose) is inoculated at 20.000 to 50.000 cells/ml. Cultures are grown at 30°C while shaking at 250 rpm until a final PD of 0.24 (+/- 0.04) 6nM is reached.

200 $\mu$l of this yeast suspension is added to all wells of MW96 plates containing R-compounds in a 450

μl total volume. MW96 plates are incubated (static) at 30°C for 48 hours.

Optical densities are measured after 48 hours.

Test growth is expressed as a percentage of positive control growth for both mutant (x) and wild type (Y) strains. The ratio (x/y) of these derived variables is calculated.

**Results**

A C. albicans integrative vector, pGAL1PSiST-1, was constructed to allow non-directional cloning of C. albicans genomic DNA fragments (Figure 2). The vector contains an inducible GAL1 promoter, a SfiI-cloned stuffer fragment, a C. albicans URA3 selection marker and elements to allow autonomous replication and selection in E coli. A C. albicans genomic DNA library was prepared by ligating small genomic DNA fragments (400 to 1000 bp) which were linked to SfiI adaptors into the SfiI opened vector pGAL1PSiST-1 vector. Genomic DNA fragments (450 ng) were ligated into the pGAL1PSiST-1 vector (20 ng). After electroporation into E. coli approximately 400,000 clones were obtained. Plasmid DNA was prepared of … clones; 91% contained an insert with an average length of 600 bp. The size of the library corresponds to over 5 times the diploid genome with genomic DNA inserts oriented in sense or antisense direction in the vector.

A similar C. albicans integrative vector, pGAL1PNiST-1, was constructed to allow SfiI/Not I directional cloning of C. albicans cDNA fragments (Figure 2). The SfiI/NotI cDNA was purified and sized on a Biogel column A150M. The first fraction contained approximately 38,720 clones upon transformation to E. coli with an average insert size of 1500 bp. cDNA from this fraction was ligated into a NotI/SfiI opened pGAL1PNiST-1 vector.

C. albicans strain CAI-4 was transformed with the aforementioned genomic and cDNA libraries. Upon homologous recombination between the insert (partial or complete gene) in a library clone and the corresponding gene in the Candida genome, this gene is (partially if the gene is not full-length) duplicated (Figure 1). The first copy of the gene is flanked upstream by its native promoter and downstream by the GAL1 promoter. The direction of transcription from the native promoter is opposite to that of the GAL1 promoter. Induction of the GAL1 promoter might thus lead to altered expression of the gene at the integration site. Moreover, if the cDNA does not contain the entire 5' coding region, the first copy of the gene may not give rise to any more to a functional protein. The second copy of this gene has lost its promoter and will therefore not be transcribed (Figure 1).

Upon integration at the site of the GAL1 promoter, the promoter is duplicated yielding an integrated gene fragment under control of the GAL1 promoter (Figure 1).

Growth curves were measured in the presence of lithium acetate. Figure 3 shows growth curves of the wild type CAI-4 strain and transformants -resulting from cDNA library transformation- showing either an overall impaired growth (clone 38; Figure 3C) or galactose-induced (clone 36; Figure 3B) reduction in growth. This analysis was performed in S-glucose medium as a non-induction medium and S-galactose medium as an induction medium. The results shown in Figure 3A show that also the wild type strain shows reduced growth in antisense induction medium. This is because galactose is used rather inefficiently as a carbon source by C. albicans. In order to solve this problem and facilitate the selection procedure an extra carbon source, maltose, was added to both inducing and non-inducing medium. Again growth patterns varied significantly from transformant to transformant but growth of the parental strain CAI-4

was nearly identical in both media (Figure 3D). Strains impaired in growth upon promoter activation showed a clear shift in the growth curve in medium supplemented with both galactose and maltose (clone 415; Figure 3E). Overall impaired growth was, as expected, not strongly influenced by the addition of maltose (clone 360; Figure 3F).

To verify that maltose as an extra carbon source did not affect the strength and inducibility of the GAL1 promoter, promoter activity was measured using Kluyveromyces lactis LAC4 reporter gene expression. CAI-4 was transformed with LAC4/pGAL1pNiST-1. Four individual transformants (named Q, R, S, T) were grown in glucose, galactose, glucose+maltose and galactose+maltose media and ß-galactosidase activity was measured (Figure 4). It is clear that the presence of maltose does not significantly influence the induction ratio of the GAL1 promoter.

From a total of over 2000 transformants screened, 198 (~10%) showed an impaired growth phenotype and were selected for further analysis. Fourty-three % of these slow growers showed a growth pattern corresponding with a putative promoter interference or antisense effect, 57% showed overall impaired growth. PCR analysis with 5pGALNiSTPCR and 3pGALNiSTPCR primers on genomic DNA from the transformants can reveal integration outside the gene showing sequence identity with the insert DNA, eg. at the GAL1 promoter region (Figure 1). Of all transformants screened by PCR using these primers, ~ 11% showed integration at a non-insert location.

When the insert of an antisense library clone recombines with the homologous gene in the C. albicans genome, no PCR product can be obtained upon amplification with 5pGALNiSTPCR and 3pGALNiSTPCR primers on genomic DNA (Figure 1). To release the plasmid from the genome and determine the integration site, genomic DNA was isolated from the transformants, cut (with SacI

for the genomic library transformants and with PstI for the cDNA library transformants), religated and the resulting ligation reaction was precipitated and used as a template for inverse PCR. This procedure reveals homologous integration at the insert site as well as the number of integrations (assuming PCR products are of different lengths) within the Candida genome. This analysis was performed on all selected transformants, ~32 % of which showed multiple integrations. The frequency of multiple integrations was very variable and depended on the batch of transformants analyzed.
The resulting PCR products from both analyses were subsequently sequenced and the sequences by compared with both public and proprietary sequence databases. In total 86 different genes could be identified, 45 of which were of unknown function.

For the CAI-4 transformants obtained with a genomic (non-directionally cloned) library, 26% of the selected clones (n=~150) contained the C. albicans autonomous replicating sequence, ARS2, and 15% of the clones contained a ribosomal RNA fragment.

For the CAI-4 transformants obtained with a cDNA library (n=~1850) a whole series of different gene fragments was found. As expected, also a number of genes involved in carbon source metabolism and nutrient uptake were identified.

Two examples of identified genes will be discussed, although as seen in Figures 16 to 70 similiar results were obtained for all of the sequences according to the invention. Clone 36 shows a galactose-induced impairment in growth, suggestive of a promoter interference or antisense effect (Figure 3B). In this clone recombination had occurred at the insert site as shown by amplification of a ~600bp gene fragment by inverse PCR. The sequence of the isolated gene fragment was 74 % identical to a S. cerevisiae S-adenosyl methionine synthetase 2 (SAM2) gene. Effects on SAM2 mRNA were

assessed by Northern blots on total RNA extracted from a non-transformed control strain and from clone 36 grown either in antisense-inducing or non-inducing media. The Northern blot was hybridised with an in vitro synthesized SAM2 RNA sense probe to detect antisense transcripts (Figure 5). An identical Northern blot was hybridised with an in vitro synthesized SAM2 antisense probe to detect SAM2 mRNA (Figure 5). Both blots were subsequently hybridized with a TEF3 DNA probe to allow normalization. As the sequence of the C. albicans SAM2 gene was not available at the time, a C. albicans genomic library in pYCp50 was prepared and E. coli transformants were screened for the full-length gene using the 600 bp SAM2 PCR fragment as a probe. A strongly hybridizing clone was identified and designated clone 36.13.1. This clone contained the complete ORF (1155 bp) of the SAM2 gene including 5′ and 3′ flanking regions. In the very A/T-rich 5′ flanking region a putative TATA box could be identified at -27 bp. The 3′ flanking region contains multiple T-rich (>10 bp) regions, elements described in yeast as necessary for transcript release (Reeder and Lang, 1997). The complete SAM2 mRNA transcript thus has a predicted length of 1.2 kb.

Clone 38 showed impaired growth in both non-inducing and inducing media (Figure 3); this is expected when integration of the library plasmid itself leads to gene suppression. Clone 38 contained a 340 bp fragment of the ribonucleotide reductase 1 (RNR1) gene. RNR1 mRNA levels were analysed by Northern blot and quantitative PCR in a non-transformed control strain and clone 38 grown in S+glucose medium. The Northern blot was hybridised successively with an actin and an RNR1 doublestranded DNA probe (Figure 6). Although the ß-actin transcript level in the control strain is lower compared to clone 38, the relative amount of RNR1 transcript is higher, indicating a reduced level of RNR1

transcript in clone 38. This result was confirmed by Taqman quantitative PCR on both control strain and clone 38 using a RNR1 fluorogenic probe. Resulting Ct values were calculated and normalised for ß-actin (Figure 6). Again RNR1 transcript levels in clone 38 were found reduced compared to the control strain.

To verify that the growth-effect was due to the interference with the identified gene and to support the spcificity of the antisense effect, single allele knock-outs were made in 6 identified genes using the URA-blaster method (Fonzi and irwin, 1993). Disruption of one allele of a gene should in theory lead to ~ 50 % reduction in gene transcript. In practice however we have observed reductions varying between 10 and 100 % of normal level. This can probably be explained by the fact that not always both copies of a gene are functional. That only a single integration at the corerct site had occurred for each of the disruption cassettes was verified by PCR and Southern blot analysis. Growth curves were measured; three disruptants showed impaired growth, suggesting that a gene required for growth or survival was targeted. Experiments to take over control of the second allele of each gene -by promoter replacement- are ongoing.

The present application describes new methods to diminish endogenous gene expression in the medically important yeast C. albicans. Our approach proved very useful for the identification of genes required for growth or survival. Technical hurdles consisted of the lack of an efficient transformation method for C. albicans (Logghe M., submitted) and the need to measure growth reproducibly on a large number of transformants in parallel. The latter was achieved with a Bioscreen Analyzer (Labsystems) which can simultaneously measure growth in 200 cultures and subsequently generate growth curves automatically. Although in principle this kind of screening could be done on plates we could not

achieve satisfactory reproducibility using plate screening.

In our genomic screen, integration of the library plasmid can happen either at the endogenous GAL1 promoter locus or, more frequently, at the locus corresponding to the plasmid insert. The latter results in a gene duplication with the first copy of the gene flanked by two convergently oriented promoters. The use of such a "collision construct" has previously been described in screening for inhibitors of transcriptional activation in mammalian cells (patent WO 97/10360; Giese K.). If RNA polymerase II complexes start from both the upstream and downstream, oppositely oriented, promoter regions, they may collide thereby preventing the formation of a full-length mRNA transcript. The second copy of the gene has no more a promoter and is probably 5' crippled as the original inserts cloned into the library have an average length of ~1.5 kb while ORFs in C. albicans have an average length of … and we ourselves identified ORFs of (unknown) genes larger than 7 kb.

Upon integration of a plasmid into the C. albicans genome, reduced function of the protein encoded by the disrupted gene can be due to several mechanisms: 1) The first copy of the duplicated gene can be prevented from forming functional sense transcript by promoter collision or the sense transcript may be inhibited by true antisense. Indeed, although a 1.2 kb SAM2 antisense transcript could be detected in clone 36 we cannot exclude the growth defect being due to promoter interference. If an antisense transcript is formed from an intact SAM2 gene, we expect a transcript of at least 1055 bp; no transcription terminator was engineered upstream of this gene so transcription will proceed until an appropriate transcription termination recognition site is reached. The promoter region of the SAM2 gene is particularly A/T rich and contains a reversed yeast transcription terminator site at - 118

(with translation starting at +1). In yeast, transcription terminator sites are ill-defined but for a T-rich stretch with non-T residues situated appropriately to prevent slippage (Jeong et al., 1996; Reeder and Lang, 1997). If termination of transcription occurs at this theoretically predicted site, a 1.17 kb transcript would be expected, as was found. 2) If mutations were present in the original library clone, the protein encoded by the gene after homologous recombination could be non-functional. 3) Possible cis down-regulatory effects on adjacent genes could be induced upon integration of large DNA fragments at certain locations within the genome. 4) Finally, gene disruption could occur by recombination with cDNA that is not full-length at the 5' end.

If -on the contrary- integration happens at the endogenous GAL1 promoter site, the GAL1 promoter is duplicated and antisense can be induced from this promoter. Promoter collision is not possible as the endogenous promoter of the gene is lacking at the integration site. Integration at a non-homologous site within the genome is rare. Transformation efficiencies of 0.7-3 transformants/µg have been reported upon transformation of CAI-4 with non-homologous plasmid DNA (Thompson et al., 1998). Also, integration at the URA3 locus is very unlikely as the complete URA3 gene has been removed from the CAI-4 genome.

Irrespective of the mechanism responsible for gene suppression, we could identify genes required for growth or survival by screening for transformants showing either galactose-induced or complete growth block. Furthermore, for such genome-wide screening no prior sequence information is needed and it allows discovery of possibly new critical functions. However, some genes may only be critical under conditions different from growth in minimal medium (as used in our screening) e.g. environments with high oxygen tension, high osmolarity

or high pH. However, it would be possible to screen for a growth phenotype under these conditions using our screening method. Alternatively, some genes may play critical roles only under unusual growth states or may specifically be required for adaptation to conditions encountered during infection of a host.

More than half of the ORFs we have identified as being critical for growth have a completely unknown function. Even though required for growth in C. albicans, for some genes no homologues could be found in existing databases, suggesting that they are species-specific genes. Indeed, recent genome sequencing efforts (e.g. Mycoplasma genitalium (Fraser et al., 1995), Haemophilus influenzae (Fleischmann et al, 1995)) have shown that approximately 20 % of the predicted ORFs in a microbial genome can be species-specific.

One disadvantage of the technique is that multiple library plasmids can integrate in the genome of a single C. albicans cell. When this occurs, identification of the target responsible for the growth defect becomes more difficult. Also, as shown in Schizosaccharomyces pombe (Atkins et al., 1995), RNA-mediated suppression may not be effective for certain genes, which we would miss when relying only on this technique.

Rather unexpectedly, transformation with the genomic library and subsequent screening for transformants showing reduced growth frequently yielded ARS2- and rRNA-containing clones (in 26 and 15% respectively of the transformants showing reduced growth). Previously, a study of aging yeast mother cells had shown that accumulation of extrachromosomal rDNA circles (ERCs) occurs in old cells and that these ERCs actually cause aging (Sinclair et al., 1997; Johnson et al., 1999). rDNA is present at 100-200 tandem copies on chromosome XII of S. cerevisiae and was found to accumulate to about 1000 copies in senescent cells. One other gene we identified is a homologue of the

essential S. cerevisiae gene TRA1, a protein kinase showing highest identity to the human TRRAP gene (McMahon et al., 1998) which is an ataxia telangiectasia mutated (ATM)-related gene. Loss of ATM is a genetic defect identified in ataxia telangiectasia (Johnson et al., 1999), a disease in humans where life span is typically reduced to 40-50 years. We might thus have picked up a number of growth-inhibitory effects due to induction of aging.

The strategy presented should be applicable to all organisms for which existing techniques for "en masse" gene disruption are not easily applicable because of their diploid nature and lack of sexual cycle and might prove especially useful for other diploid imperfect yeasts.

Although the genomic strategy that we described cannot substitute for a comprehensive investigation of individual genes and pathways, it can provide a good starting point for such investigation.

**References**

1. Altschul, S.F., W. Gish, W. Miller, E.W. Myers, and D.J. Lipman. 1990. Basic local alignment search tool. J. Mol. Biol. 215: 403-410.

2. Arndt, G.M., D. Atkins, M. Patrikakis, and J.G. Izant 1995. Gene regulation by antisense RNA in the fission yeast Schizosaccharomyces pombe. Mol. Gen. Genet. 248: 293-300.

3. Atkins, D., and W.L. Gerlach. 1994. Artificial ribozyme and antisense gene expression in S. cerevisiae. Antisense research and development 4:109-117.

4. Atkins, D., G.M. Arndt, and J.G. Izant 1994. Antisense gene expression in yeast. Biol. Chem. Hoppe-Seyler. 375:721-729.

5. Atkins, D., M. Patrikakis, J.G. Izant 1995. The ade6 gene of the fission yeast as a target for antisense and ribozyme RNA-mediated suppression. Antisense Research & Development. 5(4):295-305.

6. Bairoch, A., and R. Apweiler. 1998. The SWISS-PROT protein sequence data bank and its supplement TrEMBL in 1998. Nucleic Acids Res. 26: 38-42.

7. Baudin, A., O. Ozier-Kalogeropoulos, A. Denouel, F. Lacroute, C. Cullin. 1993. A simple and efficient method for direct gene deletion in Saccharomyces cerevisiae. Nucleic Acids Research. 21(14):3329-30.

8. Blin, N., and D.W. Stafford. 1976. Nucleic Acids Res. 3: 2303-2308.

9. Dujon, B. 1998. European Functional Analysis Network (EUROFAN) and the functional analysis of the Saccharomyces cerevisiae genome. Electrophoresis. 19:617-624.

10. Del Rosario, M., J.C. Stephans, J. Zakel, J. Escobedo, and K. Giese. 1996. Positive selection system to screen for inhibitors of human immunodeficiency virus-1 transcription. Nature Biotechnology. 14:1592-1596.

11. Fairhead, C., A. Thierry, F. Denis, M. Eck, and B. Dujon. 1998. "Mass-murder" of ORFs from three regions of chromosome XI from Saccharomyces cerevisiae. Gene. 223:33-46.

12. Ferbeyre, G., J. Bratty, H. Chen, R. Cedergren. 1996. Cell cycle arrest promotes trans-hammerhead ribozyme action in yeast. Journal of Biological Chemistry. 271(32):19318-23.

13. Fleischmann, R.D., M.D. Adams, O. White, R.A. Clayton, E.F. Kirkness, A.R. Kerlavage , C.J. Bult, J.F. Tomb, B.A. Dougherty, J.M. Merrick. 1995. Whole-genome random sequencing and assembly of Haemophilus influenzae. Science. 269: 496-512

14. Fonzi, W.A., and M.Y. Irwin. 1993. Isogenic strain construction and gene mapping in Candida albicans. Genetics 134:717-728.

15. Fraser, C.M., J.D. Gocayne, O. White, M.D. Adams, R.A. Clayton, R.D. Fleischmann, C.J. Bult, A.R. Kervalage, G. Sutton, J.M. Kelley. 1995. The minimal gene complement of Mycoplasma genitalium. Science 270:397-403.

16. Giese, K. 1997. Method and construct for screening for inhibitors of transcriptional activation. International patent application WO 97/10360.

17. Hahn, S. Guarente L. 1988. Yeast HAP2 and HAP3: transcriptional activators in a heteromeric complex. Science. 240(4850):317-21

18. Heid, C.A., J. Stevens, K.J. Livak, and P.M. Williams. 1996. Real time quantitative PCR. Genome Methods 6:986-994.

19. Jayaram, M., A. Sutton, J.R. Broach. 1985. Properties of REP3: a cis-acting locus required for stable propagation of the Saccharomyces cerevisiae plasmid 2 microns circle. Molecular & Cellular Biology. 5(9):2466-75.

20. Jeong, S.W., W.H. Lang, R.H. Reeder. 1996. The yeast transcription terminator for RNA polymerase I is designed to prevent polymerase slippage. Journal of Biological Chemistry. 271(27):16104-10.

21. Johnson, F.B., D.A. Sinclair, and L. Guarente. 1999. Molecular Biology of aging. Cell. 96:291-302.

22. Leuker, C.E., A. Sonneborn, S. Delbruck, J.F. Ernst. 1997. Sequence and promoter regulation of the PCK1 gene encoding phosphoenolpyruvate carboxykinase of the fungal pathogen Candida albicans. Gene. 192(2):235-40.

23. Lie, Y.S., and C.J. Petropoulos. 1998. Advances in quantitative PCR technology: 5' nuclease assays. Current Opinion in Biotechnology 9:43-48.

24. Livak, K.J., S.J. Flood, J. Marmaro, W. Giusti, K. Deetz. 1995. Oligonucleotides with fluorescent dyes at opposite ends provide a quenched probe system useful for detecting PCR product and nucleic acid hybridization. Genome Research. 4(6):357-62.

25. Mandart, E. 1998. Effects of mutations in the Saccharomyces cerevisiae RNA14 gene on the abundance and polyadenylation of its transcripts. Mol. Gen. Genet. 258:16-25.

26. McMahon, S.B., H.A. Van Buskirk, K.A. Dugan, T.D. Copeland, and M.D. Cole. 1998. The novel ATM-related protein TRRAP is an essential cofactor for the c-Myc and E2F oncoproteins. Cell. 94:363-74.

27. Murray, J.A.H., M. Scarpa, N. Rossi, G. Cesareni. 1987. Antagonistic controls regulate copy number of the yeast 2μ plasmid. EMBO J. 6:4205-4212.

28. Nasr, F., A. Bécam, P.P. Slonimski, and C.J. Herbert. 1994. YBR1012, an essential gene from S. cerevisiae: construction of an RNA-antisense conditional allele and isolation of a multicopy suppressor. CR Acad. Sci. Paris. 317:607-613

29. Nasr, F., A. Bécam, S.C. Brown, D. De Nay. P.P. Slonimski, and C.J. Herbert. 1995. Artificial antisense regulation of YBR1012 an essential gene from S. cerevisiae which is important for progression through G1/S. Mol. Gen. Genet. 249:51-57.

30. Nomura, T., N. Fujita, A. Ishihama. 1985.

Promoter selectivity of E. coli RNA polymerase: analysis of the promoter system of convergently-transcribed dnaQ-rnh genes. Nucleic Acids Research. 13(21):7647-61.

31. Orlando, C., P. Pinzani, and M. Pazzagli. 1998. Developments in quantitative PCR. Clin. Chem. Lab. Med. 36(5):255-269.

32. Pla, J., C. Gil, F. Monteoliva, M. Sanchez, and C. Nombelà. 1996. Understanding Candida albicans at the molecular level. Yeast. 12:1677-1702.

33. Reeder, R.H. and W.H. Lang. 1997. Terminating transcription in eukaryotes: lessons learned from RNA polymerase I. Trends in Biochemical Sciences. 22(12):473-7, 1997

34. Sambrook, J., E.F. Fritsch, and T. Maniatis. 1989. Molecular Cloning: A Laboratory Manual, 2nd Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

35. Sinclair, D.A.,L. Guarente. 1997. Extrachromosomal rDNA circles--a cause of aging in yeast. Cell. 91(7):1033-42.

36. Smith, V., D. Botstein, and P. O. Brown. 1995. Genetic footprinting: A genomic strategy for determining a gene's function given its sequence. Proc. Natl. Acad. Sci. USA. 92:6479-6483.

37. Stoesser, G., Moseley M.A., Sleep J., McGowran M., Garcia-Pastor M., Sterk P. 1998. Nucleic Acids Res. 26(1):8-15.

38. Thompson-Jager, S. Domdey H. 1990. The intron of

the yeast actin gene contains the promoter for an antisense RNA. Current Genetics. 17(3):269-73.

39. Thompson, J.R., E. Register, J. Curotto, M. Kurtz, and R. Kelly. 1998. An improved protocol for the preparation of yeast cells for transformation by electroporation. Yeast. 14:565-571.

40. Thrash, C., A.T. Bankier, B.G. Barrell, and R. Sternglanz. 1985. Proc. Natl. Acad. Sci. USA 82: 4374-4378.

41. Van Duin, M., J. van Den Tol, J.H. Hoeijmakers, D. Bootsma, I.P. Rupp, P. Reynolds, L. Prakash, and S. Prakash. 1989. Conserved pattern of antisense overlapping transcription in the homologous human ERCC-1 and yeast RAD10 DNA repair gene regions. Molecular & Cellular Biology. 9(4):1794-8.

42. Wach, A., A. Brachat, R. Pohlmann, P. Philippsen. 1994. New heterologous modules for classical or PCR-based gene disruptions in Saccharomyces cerevisiae. Yeast. 10(13):1793-808.

43. Wilson, R.B., D. Davis, A.P. Mitchell. 1999. Rapid hypothesis testing with Candida albicans through gene disruption with short homology regions. Journal of Bacteriology. 181(6):1868-74.

44. Zhu, J., W. Kempenaers, D. Van der Straeten, R. Contreras, and W. Fiers. 1985. A method for fast and pure DNA elution from agarose gels by centrifugal filtration. Bio/Technology. 3: 1014-1016.

## TABLE 1

| Seq ID No. | Clone | Function |
|---|---|---|
| 1 | 214c_cpL1 | - |
| 2 | 113g2 | - |
| 3 | 222g8 | - |
| 4 | 222g8(prt) | - |
| 5 | 222g9 | - |
| 6 | 222g9_CDS_1 | - |
| 7 | 222g9_CDS_2 | - |
| 8 | 222g9_CDS_3 | - |
| 9 | 222g9_CDS_4 | - |
| 10 | 24gG | - |
| 11 | 28gK | - |
| 12 | 328c1 | - |
| 13 | 328c1(prt) | - |
| 14 | 33gK | - |
| 15 | 33gK(prt) | - |
| 16 | 3gG | - |
| 17 | 58gA | - |
| 18 | 21g2 | - |
| 19 | 21g2(prt) | 5' UTR TRA1 |
| 20 | 223c_cp | CFL |
| 21 | 357cL | |
| 22 | 357cL(prt) | RPL27 |
| 23 | 110c_af | |
| 24 | 110c_af(prt) | SADH |
| 25 | CDC48 | |
| 26 | CDC48(prt) | CDC48 |
| 27 | 99g3 | |
| 28 | 99g3(prt) | CIT |
| 29 | ESP1 | |
| 30 | ESP1(prt) | ESP1 |
| 31 | 190g3 | |
| 32 | 190g3(prt) | FAL1 |
| 33 | 249c_af | |
| 34 | 249c_af(prt) | MAA |
| 35 | 485cL | |
| 36 | 485cL(prt) | RPL16 |

TABLE 1 (CONTINUED)

| | | |
|---|---|---|
| 37 | 328c3 | |
| 38 | 328c3(prt) | RPS21 |
| 39 | 83c3 | |
| 40 | 83c3(prt) | SHA3 |
| 41 | 233c_cp2 | |
| 42 | 233c_cp2 | TPI1 |
| 43 | 214c_cpL1 | HXT6_2 |
| 44 | 128g4 | 15S rRNA |
| 45 | 135g | tRNA_Ser |
| 46 | 22g3 | |
| 47 | 22g3_CDS1 | |
| 48 | 22g3_CDS2 | - |
| 49 | 38g1 | - |
| 50 | 117c_af | - |
| 51 | 117c_af(prt) | - |
| 52 | 17g1 | - |
| 53 | 17g1_CDS1 | - |
| 54 | 17g1_CDS2 | - |
| 55 | 480c | - |
| 56 | 480c(prt) | - |
| 57 | 55g3 | - |
| 58 | 55g3(prt) | - |
| 59 | 61gB | |
| 60 | 61gB(prt) | PSP2 |
| 61 | 62gB | |
| 62 | 62gB(prt) | - |
| 63 | 80g3 | |
| 64 | 80g3(prt) | - |
| 65 | 29g2_part1 | |
| 66 | 29g2_part1(prt) | EF4 |
| 67 | 29g2_part2_3 | |
| 68 | 29g2_part2(prt) | EF4 |
| 69 | 29g2_part3(prt) | EF4 |
| 70 | 226c_af2 | |
| 71 | 226c_af2(prt) | ADE12 |
| 72 | 409c5 | |
| 73 | 409c5(prt) | HOL1 |
| 74 | 40c_af | |
| 75 | 40c_af(prt) | FBP |
| 76 | 55g1 | |

TABLE 1 (CONTINUED)

| | | |
|---|---|---|
| 77 | 55g1(prt) | MEG1 |
| 78 | 67g1 | |
| 79 | 67g1(prt) | RVS167 |
| 80 | 232c_cp | |
| 81 | 360c6 | |
| 82 | 360c6(prt) | HXT6_1 |
| 83 | 98c_cp | |
| 84 | 98c_cp(prt) | KGD2 |
| 85 | 17c_cp | |
| 86 | 17c_cp(prt) | NDE1 |
| 87 | 60gK | |
| 88 | 60gK(prt) | RAD18 |
| 89 | 226c_af1 | |
| 90 | 226c_af1(prt) | - |
| 91 | 328c2 | |
| 92 | 328c2(prt) | - |
| 93 | 498c_cp | |
| 94 | 498c_cp(prt) | - |
| 95 | 64gB | |
| 96 | 64gB(prt) | - |
| 97 | 8c_cp | |
| 98 | 8c_cp(prt) | - |
| 99 | 15c1 | |
| 100 | 15c1(prt) | - |
| 101 | 233c_cp1 | |
| 102 | 233c_cp1_CDS1 | |
| 103 | 233c_cp1_CDS2 | - |
| 104 | 35gK | |
| 105 | 35gK(prt) | - |
| 106 | 36g2 | |
| 107 | 36g2(prt) | - |
| 108 | 65g | |
| 109 | 65g(prt) | - |
| 110 | 85g3 | |
| 111 | 85g3(prt) | |
| 112 | 232c_cp(prt) | SAP |
| 113 | 409c10 | |
| 114 | 409c10(prt) | - |